Opportunistic Spectrum Access in Unknown Dynamic Environment: A Game-Theoretic Stochastic Learning Solution

Yuhua Xu, *Student Member, IEEE*, Jinlong Wang, *Member, IEEE*, Qihui Wu, *Member, IEEE*, Alagan Anpalagan, *Senior Member, IEEE*, and Yu-Dong Yao, *Fellow, IEEE*

Abstract—We investigate the problem of distributed channel selection using a game-theoretic stochastic learning solution in an opportunistic spectrum access (OSA) system where the channel availability statistics and the number of the secondary users are apriori unknown. We formulate the channel selection problem as a game which is proved to be an exact potential game. However, due to the lack of information about other users and the restriction that the spectrum is time-varying with unknown availability statistics, the task of achieving Nash equilibrium (NE) points of the game is challenging. Firstly, we propose a genie-aided algorithm to achieve the NE points under the assumption of perfect environment knowledge. Based on this, we investigate the achievable performance of the game in terms of system throughput and fairness. Then, we propose a stochastic learning automata (SLA) based channel selection algorithm, with which the secondary users learn from their individual action-reward history and adjust their behaviors towards a NE point. The proposed learning algorithm neither requires information exchange, nor needs prior information about the channel availability statistics and the number of secondary users. Simulation results show that the SLA based learning algorithm achieves high system throughput with good fairness.

Index Terms—Cognitive radio networks, opportunistic spectrum access, distributed channel selection, exact potential game, stochastic learning automata.

I. INTRODUCTION

OPPORTUNISTIC spectrum access (OSA), which mainly builds on the cognitive radio technology [1], has been regarded as a promising solution to lessen the spectrum scarcity problem and hence has drawn great attention [2]–[4]. Currently, although some progress with regard to game-theoretic channel selections for OSA systems has been achieved in the literature, e.g., [5]–[7], there are still several unsolved problems. First, most existing work is based on the assumption that the secondary users have full knowledge about the environment and complete information about actions taken by other users. Secondly, the environment is required to be static during the convergence of the algorithms. However, these assumptions are not realistic in practice, because (i) obtaining the environment knowledge consumes a lot network resources, e.g., time, power, and bandwidth, and may not be feasible in some scenarios; moreover, acquiring information about other users also leads to heavy communication overhead, and (ii) the spectrum environment is always time-varying in OSA systems.

In this article, we study the problem of distributed channel selection in OSA systems using a game-theoretic learning solution. The reasons for using a game model are twofold. First, the secondary users make decisions spontaneously and independently. Secondly, their objectives are conflicting and their decisions are interactive. We consider an OSA system with the following characteristics: (i) the spectrum holes are time-varying, (ii) no need to know the channel availability statistics and the number of the secondary users in the system and (iii) no need for information exchange between the secondary users. Notably, the only available information in the considered system is the individual history of each user's channel selection decisions and rewards. Thus, a learning solution that can learn from the history information and achieve Nash equilibrium (NE) points in the unknown and dynamic spectrum environment is desirable.

To achieve this goal, we formulate the distributed channel selection problem as a game which is proved to be an exact potential game. To address the lack of information about others and the time-varying spectrum environment, we propose a stochastic learning automata (SLA) [8] based algorithm, with which the secondary users learn from their individual action-reward history and adjust their behaviors towards a NE point. To summarize, the main contributions of this article are as follows:

- Under the assumption of perfect environment information, we first propose a genie-aided channel selection algorithm to achieve NE points of the game, and investigate the system performance in terms of throughput and fairness.
- To address the lack of information and time-varying spec-

trum environment, we propose a SLA based algorithm to achieve the NE points, which neither requires information exchange, nor needs prior information about the channel availability statistics and the number of secondary users.

The rest of this article is organized as follows. In Section II, we review the related work. In Section III, we present the system model as well as the problem formulation. In Section IV, we formulate the channel selection game and investigate the properties of its NE points by proposing a genie-aided algorithm. In Section V, we propose a SLA based algorithm to achieve NE points of the game. In Section VI, simulation results and discussion are presented. Finally, we make conclusion in Section VII.

II. RELATED WORK

The problem of distributed channel selection in OSA systems has drawn great attention, and several literature addressing this topic from non-game theoretic and game theoretic perspective can be found. The framework of partially observable Markov decision process based opportunistic spectrum access was first established in [9], and further investigated in [10]–[12]. A main drawback of those studies is that they did not consider the interactions among multiple secondary users.

The problem of distributed channel selection with unknown channel availability statistics, which is the focus of this article, begins to draw attention in [13]–[16] using the multi-armed bandit approaches. There are two key differences in our work. First, their optimization objectives are minimizing the regret, whereas in our work the objective is maximizing the system aggregate throughput. Secondly, those authors mainly focused on investigating the asymptotical performance of their learning algorithms and balancing the tradeoff between exploration and exploitation. We study this problem from a game-theoretic perspective, and focus on studying the properties of the game as well as investigating a learning algorithm that converges towards NE in the unknown and dynamic spectrum environment.

It is known that learning is the core in cognitive radio [2]. In [17], a multi-agent Q-learning channel selection algorithm for the scenario of two users and two channels was proposed. However, the considered system model therein is static and the proposed learning algorithm is limited to two-by-two case. Moreover, the stochastic learning automata (SLA) [8], which can be regarded as an adaptive decision-making mechanism that learns the optimal action from a set of actions through repeated interactions with a random environment, has been successfully applied into wireless communication systems [18]–[21].

The convergence towards NE of SLA based algorithms is a key concern but not well studied. In [8], the authors have investigated the convergence of SLA based algorithm towards NE for coordination games with all the players having exactly the same utility function. In [21], the authors have investigated the convergence for a type of exact potential game, in which the received rewards in each iteration are the same for all the players. In this article, we investigate the convergence towards NE for a general exact potential game, in which the users have different utility functions and receive different rewards
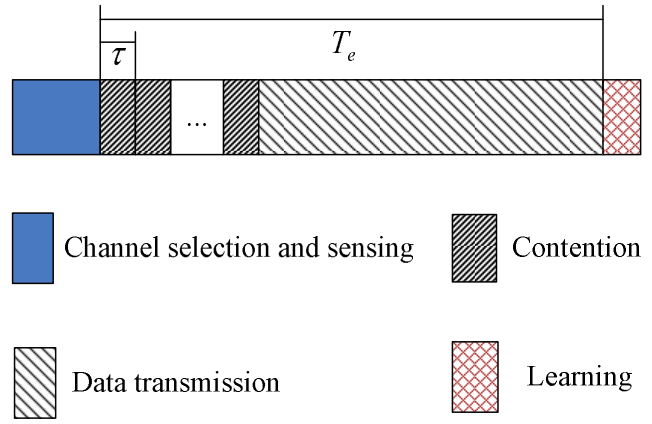

Fig. 1.   Transmission structure of the secondary users.

in each iteration. Thus, the results obtained in this article can be regarded as a significant extension of existing work and can hence be applied into more general scenarios.

III. SYSTEM MODEL AND PROBLEM FORMULATION

A. System model

We consider an opportunistic spectrum access (OSA) system involving N secondary users and M licensed channels with transmission rate R_m, $1 \leq m \leq M$, $N > M > 1$. It is assumed that each channel supports the same transmission rate for all users. This represents the case that each channel yields the same bandwidth and same transmission rate to each user, although different users may experience different channel conditions [22]. Note that such an assumption holds in some practical systems, e.g., IEEE 802.16d/e standard [23]. Moreover, it is assumed that the primary users use the licensed channels in a slotted fashion and their activities are independent from channel to channel and from slot to slot. Thus, channel m can be assumed to be idle with probability θ_m in each slot, $1 \leq m \leq M$. To make it more practical, the following are considered in this work:

1) The channel availability statistics θ_m, $1 \leq m \leq M$, are fixed but unknown.
2) The number of the secondary users, N, is unknown.
3) There is no centralized controller and no information exchange between the secondary users.

The transmission structure of the secondary users is described in Fig. 1. We assume that the secondary users can only select one channel for transmission in a slot due to hardware limitation [9]. At the beginning of each slot, each user selects a channel to sense according to its current channel selection strategy. For simplicity, it is assumed that the channel sensing is perfect[1]. If the selected channel is sensed idle, a carrier sense multiple access (CSMA) mechanism can be implemented to address collision issue among the secondary users sharing the same channel [9]. Otherwise, the secondary user has to suspend its transmission in this slot. At the end of the slot, the secondary user receives a random reward, and then applies a learning algorithm to update its channel selection strategy.

---

[1]The analysis is this article can easily be extended to the scenario with imperfect channel sensing.

In the contention process, time is divided into mini-slots with equal length and each secondary user contends for the channel access with the same probability p_a in each mini-slot[2]. A channel contention of a secondary user is said to be successful if no other secondary user contends during the same mini-slot. After a successful channel contention, the successful secondary user transmits data in the residual time slot. Once the channel is occupied by a secondary user, all other competing secondary users in the same channel have to keep silent until the next slot. For presentation, we denote the useful time after sensing in a slot as T_e, and the mini-slot length as τ (see Fig. 1).

B. Problem formulation

Let a_n denote a channel selection of secondary user n, c_m denote the set of secondary users who select channel m for transmission, i.e., $c_m = \{n \in \{1, \ldots, N\} : a_n = m\}$, and s_m denote the number of these secondary users, i.e., $s_m = |c_m|$. Then the random reward, which is equivalent to the normalized effective transmission time in a slot, received by secondary user n, $\forall n \in c_m$, in the jth slot is given by:

$$r_n(j) = \left[(T_e - N_c(s_m)\tau)/T_e\right]\beta_n(s_m)I_m R_m, \quad (1)$$

where T_e is the useful time after channel sensing in a slot, τ is the length of a mini-slot, $N_c(s_m)$ is the number of mini-slots it takes for a successful channel contention with the condition that there are s_m secondary users contending for this channel, $\beta_n(s_m)$ indicates whether secondary user n successfully contends the channel or not, and I_m indicates whether the channel m is idle or occupied. This rewarding strategy captures the dynamics of the random radio environment and the interactions of multiple secondary users for channel access. It can be seen that $N_c(s_m)$ is a geometric random variable [25] with the following probability mass function (PMF):

$$\Pr\{N_c(s_m) = i\} = p_s(1 - p_s)^{i-1}, i \geq 1, \quad (2)$$

where $p_s = s_m p_a (1 - p_a)^{s_m - 1}$ represents the overall successful channel contention probability in a mini-slot. Moreover, $\beta_n(s_m)$ and I_m are Bernoulli random variables with the following PMFs:

$$\Pr\{\beta_n(s_m) = x\} = \begin{cases} \frac{1}{s_m}, & x = 1 \\ 1 - \frac{1}{s_m}, & x = 0 \end{cases}, \quad (3)$$

and

$$\Pr\{I_m = y\} = \begin{cases} \theta_m, & y = 1 \\ 1 - \theta_m, & y = 0 \end{cases} \quad (4)$$

respectively.

Let us define a *throughput loss function* due to the contention of multiple secondary users as follows:

$$f(s_m) = \mathbf{E}[T_e - N_c(s_m)\tau]/T_e, \quad (5)$$

where $\mathbf{E}[\cdot]$ takes the expectation.

Based on (1)–(5), the expected throughput achieved by secondary user n in a slot is given by:

$$\bar{r}_n(j) = \frac{\theta_m f(s_m) R_m}{s_m}, \quad (6)$$

and the system throughput, which is defined as the aggregate throughput obtained by all the secondary users, is given by:

$$U_s(a) = \sum_{n=1}^{N} \frac{\theta_{a_n} f(s_{a_n}) R_{a_n}}{s_{a_n}} = \sum_{m=1}^{M} \theta_m f(s_m) R_m \delta(s_m), \quad (7)$$

where $a = (a_1, \ldots, a_N)$ is a channel selection profile for the secondary users, and $\delta(s_m)$ is the following indicator function:

$$\delta_m = \begin{cases} 1, s_m \geq 1 \\ 0, s_m = 0 \end{cases} \quad (8)$$

Then the system-centric objective is to find the optimal channel selection profile a_{opt} such that the system throughput is maximized. Formally,

$$a_{\text{opt}} = \arg\max U_s(a) \quad (9)$$

It is seen that the task of solving (9) is challenging, since (i) there is no control center for secondary users, and (ii) θ_m and N are unknown. Thus, a distributed approach with learning ability is desirable.

Remark 1: The secondary users can also adopt other CSMA schemes. To make it more general, the discussion in the rest of this article does not rely on the specific expression of $f(s_{a_n})$; instead, the following properties of the general CSMA schemes are used:

- $s_{a_n} \geq 1, \forall n \in \{1, \ldots, N\}$.
- In general, $f(s_{a_n})$ is a decreasing function of s_{a_n}.
- In general, $0 \leq f(s_{a_n}) \leq 1, \forall s_{a_n} \geq 1$; moreover, we have $f(s_{a_n}) \equiv 1, \forall s_{a_n} \geq 1$, when the contention overhead is negligible.

IV. GAME-THEORETIC DISTRIBUTED CHANNEL SELECTION

A. Game model

We formulate the problem of distributed channel selection in the OSA system as a game. Formally, the game is denoted by $\mathcal{G}_c = [\mathcal{N}, \{A_n\}_{n \in \mathcal{N}}, \{u_n\}_{n \in \mathcal{N}}]$, where $\mathcal{N} = \{1, \ldots, N\}$ is the set of players (secondary users), $A_n = \{1, \ldots, M\}$ is the set of available actions (licensed channels) for player n, and u_n is the utility function of player n, which is defined as the expected achievable throughput of player n, i.e.,

$$u_n(a_n, a_{-n}) \triangleq \mathbf{E}[r_n|(a_n, a_{-n})] = \frac{\theta_{a_n} f(s_{a_n}) R_{a_n}}{s_{a_n}}, \quad (10)$$

where r_n is the random reward received by player n as specified by (1), $a_n \in A_n$ represents the channel selection of player n, and $a_{-n} \in A_1 \times \cdots \times A_{n-1} \times A_{n+1} \times \cdots \times A_N$ represents a channel selection profile of all the players excluding n, where $\times$ denotes the Cartesian product.

Definition 1 (Nash equilibrium): A channel selection profile $a^* = (a_1^*, \ldots, a_N^*)$ is a pure strategy NE point of $\mathcal{G}_c$ if and only if no secondary user can improve its utility function by deviating unilaterally, i.e.,

$$\frac{\theta_{a_n^*} f(s_{a_n^*}) R_{a_n^*}}{s_{a_n^*}} \geq \frac{\theta_{a_n} f(s_{a_n}+1) R_{a_n}}{s_{a_n}+1}, \forall n \in \mathcal{N}, \forall a_n \in A_n \backslash \{a_n^*\}. \tag{11}$$

where $f()$ is the throughput loss function specified by (5), $s_{a_n^*}$ is the number of secondary users selecting channel a_n^*, and $A_n \backslash a_n$ means that a_n is excluded from A_n.

B. Analysis of NE

The properties of NE of $\mathcal{G}_c$ are characterized by the following theorems.

Theorem 1. $\mathcal{G}_c$ *is an exact potential game which has at least one pure strategy NE point.*

Proof: For a channel selection profile $a = (a_n, a_{-n})$, denote the number of secondary users selecting each channel m as $s_m, \forall m \in \{1, \ldots, M\}$. Then, we define the following potential function $\phi : A_1 \times \cdots \times A_N \to R$ for the channel selection game $\mathcal{G}_c$:

$$\phi(a_n, a_{-n}) = \sum_{m=1}^{M} \sum_{k=1}^{s_m} \varphi_m(k), \tag{12}$$

where $\varphi_m(k) \triangleq \frac{\theta_m f(k) R_m}{k}$. The above function is also known as *Rosenthal's potential function* [26].

Suppose that an arbitrary player n unilaterally changes its channel selection from a_n to $\tilde{a}_n$, then the change in individual utility function caused by this unilateral change is given by:

$$u_n(\tilde{a}_n, a_{-n}) - u_n(a_n, a_{-n}) = \varphi_{\tilde{a}_n}(s_{\tilde{a}_n}+1) - \varphi_{a_n}(s_{a_n}). \tag{13}$$

Since player n's channel selection change only affects the users on channels a_n and $\tilde{a}_n$, the change in the potential function caused by this unilateral change is given by:

$$\begin{aligned}
& \Phi(\tilde{a}_n, a_{-n}) - \Phi(a_n, a_{-n}) \\
&= \left(\sum_{k=1}^{s_{\tilde{a}_n}+1} \varphi_{\tilde{a}_n}(k) + \sum_{k=1}^{s_{a_n}-1} \varphi_{a_n}(k) \right) \\
& \quad - \left(\sum_{k=1}^{s_{\tilde{a}_n}} \varphi_{\tilde{a}_n}(k) + \sum_{k=1}^{s_{a_n}} \varphi_{a_n}(k) \right) \\
&= \varphi_{\tilde{a}_n}(s_{\tilde{a}_n}+1) - \varphi_{a_n}(s_{a_n}).
\end{aligned} \tag{14}$$

From (13) and (14), we have the following equation:

$$u_n(\tilde{a}_n, a_{-n}) - u_n(a_n, a_{-n}) = \Phi(\tilde{a}_n, a_{-n}) - \Phi(a_n, a_{-n}). \tag{15}$$

It is seen from (15) that the change in individual utility function caused by any player's unilateral deviation is the same as the change in the potential function. Thus, according to the definition given in [27], the channel selection game $\mathcal{G}_c$ is an exact potential game with potential function ϕ. Exact potential game belongs to potential games, which exhibit several nice properties and the most important one is that every potential game has at least one pure strategy NE point. Therefore, Theorem 1 is proved. ∎

Although Theorem 1 shows that the proposed game $\mathcal{G}_c$ has at least one pure strategy NE, the NE points can not be straightforwardly obtained. In order to investigate the properties of the game, we assume that there is an omnipotent

---

Algorithm 1: Genie-aided channel selection algorithm

1) Initially, set $\mathcal{N}_1 = \mathcal{N}$, $\mathcal{N}_2 = \emptyset$, $k = 1$ and $s_m(k) = 0$, $1 \leq m \leq M$, where $\emptyset$ is the null set.
2) Randomly select a secondary user $n \in \mathcal{N}_1$ and let it select channel m^*, i.e., $a_n = m^*$, where m^* is determined by:

$$m^* \in \underset{1 \leq m \leq M}{\arg \max} \left[\frac{\theta_m f(s_m(k)+1) R_m}{s_m(k)+1} \right] \tag{16}$$

That is, m^* is the one which leads to the maximum individual throughput.
3) Exclude n from $\mathcal{N}_1$ and include it in $\mathcal{N}_2$, i.e., $\mathcal{N}_1 = \mathcal{N}_1 \backslash n$ and $\mathcal{N}_2 = \mathcal{N}_2 \cup n$. Then, update $\{s_1(k), \ldots, s_M(k)\}$ according to the following rules:

$$\begin{aligned}
s_m(k+1) &= s_m(k)+1, \quad m = m^* \\
s_m(k+1) &= s_m(k), \qquad m \neq m^*
\end{aligned} \tag{17}$$

4) If $\mathcal{N}_1 = \emptyset$, stop; else go to step 2).

---

genie, which knows the channel availability statistics and can perfectly monitor all selections made by the secondary users in each iteration. Then, we propose a genie-aided channel selection algorithm, which is described by Algorithm 1, to find NE points of $\mathcal{G}_c$.

Theorem 2. *The proposed genie-aided selection algorithm converges to a pure strategy NE point of $\mathcal{G}_c$.*

Proof: Refer to Appendix A. ∎

It is noted that in Step 2) of the genie-aided algorithm, there may be multiple channels simultaneously resulting in the maximum individual throughput for the selected user. That is, these channels are indistinguishable for the user. In this case, although the algorithm still achieves a NE point of the game, the final number of secondary users on the channels, i.e., $\{s_1(N), \ldots, s_M(N)\}$, is not deterministic since it depends on which channel the user selects among those indistinguishable channels. For general scenarios, it is hard to investigate the achievable system throughput of the NE solutions. However, we can investigate it under the assumption (it is tagged as **A1**) that the channels are distinguishable in each iteration, which provides a better understanding of this problem.

Theorem 3. *Under assumption A1, the channel selection game $\mathcal{G}_c$ has multiple pure strategy NE points, and all of them lead to the same system throughput.*

Proof: It can be seen from the procedure of the genie-aided channel selection algorithm that if the assumption A1 holds, it will finally lead to a deterministic and unique channel selection scheme $\{s_1^*(N), \ldots, s_M^*(N)\}$. Moreover, since the secondary users are randomly selected in each iteration, the unique channel selection scheme corresponds to multiple channel selection profiles. Because it does not matter which secondary user selects the channel, but how many secondary users select the channel. Thus, according to Theorem 2, there are multiple pure strategy NE points in $\mathcal{G}_c$. More specifically,

the number of the pure strategy NE points is as follows:

$$K = \binom{N}{s_1^*(N)} \binom{N - s_1^*(N)}{s_2^*(N)} \cdots \binom{N - \sum_{m=1}^{M-2} s_m^*(N)}{s_{M-1}^*(N)} \quad (18)$$

Then, according to the definition of system throughput specified by (7), Theorem 3 follows. ∎

Besides the system throughput, fairness is another concern in wireless communication systems. Let us present an example to illustrate this. Suppose that there is an OSA system with two channels and three secondary users. The channels are with unit transmission rate and the channel idle probabilities are 0.7 and 0.6 respectively. It is known that the NE solution in this case is as follows: two users select the channel with idle probability 0.7, while the other selects the one with 0.6. Without considering the throughput loss due to contention, the individual throughput for the users are $(0.35\ 0.35\ 0.6)$, which is relatively unfair. Thus, we conclude that due to the selection of different channels, the obtained throughput may vary from user to user and hence the fairness in the channel selection game needs to be investigated.

In the following, we investigate the fairness of $\mathcal{G}_c$ using Jain's fairness index (JFI) [28]. Formally, the JFI of $\mathcal{G}_c$ is defined as follows:

$$J_{\mathcal{G}_c} = \frac{\left(\sum_{n=1}^{N} u_n\right)^2}{N \sum_{n=1}^{N} u_n^2}, \quad (19)$$

where $u_n \geq 0$, $\forall n$, denotes the expected throughput obtained by the secondary user n. JFI translates a resource allocation vector $\{u_1, \ldots, u_N\}$ into a score in the interval of $[1/N, 1]$. It is noted that higher JFI implies that the resource allocation is more fair. Specifically, $J_{\mathcal{G}_c} = 1$ corresponds to an absolute fair system (in this case, all the users get the same amount of resource, i.e., $u_n = u_0, \forall n$), and $J_{\mathcal{G}_c} = 1/N$ corresponds to an absolute unfair system (in this case, there exists a user such that $u_n > 0$ and $u_k = 0, \forall k \neq n$). The following theorem characterizes the achieved fairness of $\mathcal{G}_c$.

Theorem 4. *If the channel contention is negligible, $J_{\mathcal{G}_c}$ is no less than $8/9$.*

Proof: Refer to Appendix B. ∎

Remark 2: Although the channel contention overhead can not be ignored in general scenarios, it is believed that the game-theoretic solution can still achieve good fairness. This is due to the nature of Nash equilibrium, which can be essentially interpreted as another form of fairness. Furthermore, simulation results presented in Section VI validate this statement.

V. STOCHASTIC LEARNING SOLUTION FOR ACHIEVING NE IN UNKNOWN DYNAMIC ENVIRONMENT

It is known that there are some learning algorithms converging towards pure strategy NE points for an exact potential game, such as regret learning [5], best (better) response dynamic [27], spatial adaptive play [22], [29] and fictitious play [30]. However, these algorithms require complete information about the actions selected by all other players in each iteration;

Algorithm 2: SLA based channel selection algorithm

---

1) Initially, set $j = 0$ and the initial channel selection probability vector $p_{nm}(j) = 1/M, \forall n \in \mathcal{N}, m \in \{1, \ldots, M\}$.
2) At the beginning of the jth slot, each secondary user n selects a channel $a_n(j)$ according to its current channel selection probability vector $\mathbf{p}_n(j)$.
3) In each slot, the secondary users perform channel sensing and channel contention. At the end of the jth slot, each secondary user n receives the random reward $r_n(j)$ specified by (1).
4) All the secondary users update their channel selection probability vectors according to the following rule:

$$p_{nm}(j + 1) = p_{nm}(j) + b\tilde{r}_n(j)(1 - p_{nm}(j)), m = a_n(j)$$
$$p_{nm}(j + 1) = p_{nm}(j) - b\tilde{r}_n(j)p_{nm}(j), \qquad m \neq a_n(j)$$
$$\qquad\qquad (21)$$

where $0 < b < 1$ is the step size, $\tilde{r}_n(j)$ is the normalized reward defined as follows:

$$\tilde{r}_n(j) = r_n(j)/(\max_m R_m) = r_n(j)/R_{\max} \quad (22)$$

5) If $\forall n \in \mathcal{N}$, there exists a component of $p_n(j)$ which is approaching one, e.g., larger than 0.99, stop; Otherwise, go to step 2).

---

in addition, the environment is required to be static during the convergence. Thus, these algorithms are not feasible in the considered OSA system. Motivated by the underlying idea of stochastic learning automata (SLA) [8], we propose a SLA based channel selection algorithm, with which the secondary users learn from their individual action-reward experiences and finally adjust their behaviors towards a NE point.

A. Algorithm description

To characterize the SLA based channel selection algorithm, we extend the channel selection game $\mathcal{G}_c$ to a mixed strategy form and give the following definitions. Let $\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$ denote the mixed strategy profile of $\mathcal{G}_c$; more specifically, $\mathbf{p}_n = (p_{n1}, \ldots, p_{nM})$, $\forall n \in \mathcal{N}$, is the channel selection probability vector of secondary user n, where p_{nm} denotes the probability with which secondary user n selects channel m. Let $h_{nm}(\mathbf{P})$ denote the expected reward function of player n if it employs pure strategy m (i.e., $a_n = m$) and other secondary users $k, \forall k \in \mathcal{N}, k \neq n$, employ mixed strategy $\mathbf{p}_k$. Formally,

$$h_{nm}(\mathbf{P}) = \sum_{a_k, k \neq n} u_n(a_1, ..., a_{n-1}, m, a_{n+1}..., a_N) \prod_{k \neq n} p_{ka_k}$$
$$\qquad\qquad (20)$$

The proposed SLA based channel selection algorithm is described in Algorithm 2. Notably, the proposed SLA based algorithm is characterized by the following two distinctive features: (i) the channel selection is based on a probability distribution over the channel set, and (ii) the probability distribution is updated based on the received reward in the competitive environment. Specifically, $r_n(j)$ serves as a reinforcement signal. If a channel is selected and the system

feeds back a positive reward, i.e., $r_n(j) > 0$, the probability of selecting this channel in the next slot increases. On the other hand, if the system feeds back zero reward, i.e., $r_n(j) = 0$, the probability of selecting this channel in the next slot remains unchanged.

Moreover, the proposed learning solution is completely distributed, since the updating rules specified by (21) is only dependent on their individual action-reward experiences. It is also noted that it neither needs any information exchange, nor monitors the actions taken by other users.

B. Discussion on convergence towards NE

In [8], the authors have investigated the convergence of the SLA based algorithms towards NE points for coordination games, which is a special kind of game in which all the players have the same utility function. Recently, the authors in [21] have proved the convergence towards NE points of a SLA based algorithm, under an assumption that all the players receive the same reward in each iteration. It is seen from (1) and (10) that the considered scenario in this article is more general since the users have different utility functions and receive different rewards in each iteration. Therefore, the proposed channel selection game $\mathcal{G}_c$ is beyond the scope of existing work, and hence the convergence of the proposed learning solution needs to be re-investigated.

The proof of the convergence is structured as follows. First, using the ordinary differential equation (ODE), we characterize the long-term behavior of the sequence $\{\mathbf{P}(j)\}$ and exploit the relationship between the stable points of the ODE and the Nash equilibria of $\mathcal{G}_c$. Secondly, we establish a sufficient condition to achieve NE points for the proposed learning solution and prove that $\mathcal{G}_c$ satisfies this condition.

Proposition 1. *With a sufficiently small step size b, the sequence $\{\mathbf{P}(j)\}$ will converge to $\mathbf{P}^*$, which is the solution of the following ODE:*

$$\frac{d\mathbf{P}}{dt} = F(\mathbf{P}), \mathbf{P}_0 = \mathbf{P}(0), \tag{23}$$

where $\mathbf{P}_0$ is the initial channel selection probability matrix and $F(\mathbf{P})$ is the conditional expected function defined as:

$$F(\mathbf{P}) = \mathbf{E}[G(\mathbf{P}(j), \mathbf{a}(j), \mathbf{r}(j))|\mathbf{P}(j)] \tag{24}$$

In (24), $G(\mathbf{P}(j), \mathbf{a}(j), \mathbf{r}(j)) = \mathbf{P}(j+1)$ represents the updating rules specified by (21).

Proof: Refer to Theorem 3.1 in [8]. ∎

Proposition 2. *The following are true of the SLA based algorithm:*
1) *All the stable stationary points of (23) are the Nash equilibria of $\mathcal{G}_c$.*
2) *All the Nash equilibria of $\mathcal{G}_c$ are the stable stationary points of (23).*

Proof: Refer to Theorem 3.2 in [8]. ∎

Theorem 5. *Suppose that there is a non-negative function $H(\mathbf{P}): \mathbf{P} \to R$ for some positive constant c such that:*

$$\begin{aligned} &H(m_1, \mathbf{P}_{-n}) - H(m_2, \mathbf{P}_{-n}) \\ &= c[h_{nm_1}(\mathbf{P}) - h_{nm_2}(\mathbf{P})], \forall n, m_1, m_2, \mathbf{P}, \end{aligned} \tag{25}$$

where $H(m, \mathbf{P}_{-n})$ is the value of H on the condition that $\mathbf{p}_n$ is a unit vector with the mth component unity, and $h_{nk}(\mathbf{P})$ is specified by (20). Then, the SLA based algorithm converges to a pure strategy NE point of a game.

Proof: Refer to Appendix C. ∎

Theorem 5 establishes a sufficient condition that can guarantee the convergence towards NE. Next, we prove that $\mathcal{G}_c$ satisfies this condition and hence it converges to a pure strategy NE point by using the SLA based learning algorithm.

Theorem 6. *With a sufficiently small step size b, the proposed SLA based channel selection algorithm converges to a pure NE point of $\mathcal{G}_c$.*

Proof: Take $H(\mathbf{P}) = \mathbf{E}[\Phi(\mathbf{P})]$, where Φ is the potential function specified by (12). Then, we have:

$$H(m, \mathbf{P}_{-n}) = \sum_{a_k, k \neq n} \Phi(a_1, ..., a_{n-1}, m, a_{n+1}, ..., a_N) \prod_{k \neq n} p_{k a_k} \tag{26}$$

Applying (15) and (20), it is easy to have:

$$H(m_1, \mathbf{P}_{-n}) - H(m_2, \mathbf{P}_{-n}) = h_{nm_1}(\mathbf{P}) - h_{nm_2}(\mathbf{P}) \tag{27}$$

By Theorem 5, Theorem 6 follows. ∎

In fact, if we take $H_e(\mathbf{P}) = \mathbf{E}[\Phi_e(\mathbf{P})]$, where $\Phi_e(\cdot)$ is the potential function of any exact potential game, then the sufficient condition specified by (25) is always met. Thus, we claim that the SLA based algorithm converges to a pure strategy NE point for any exact potential game.

Remark 3: It is noted from Theorem 6 that when b approaches zero, the proposed SLA based learning algorithm finally converges to a NE point. However, smaller step size b implies a slower convergence speed. Hence, the choice of the step size b involves a tradeoff between accuracy and speed, and is application-dependent. This can be done by practical experiments or training [20], [21].

VI. SIMULATION RESULTS AND DISCUSSION

In this section, we simulate the convergence and the throughput performance of the SLA based channel selection algorithm. The slot length is set to $T = 100 \times 10^{-3}$s; moreover, to meet the sensing requirement, the sensing length in a slot is set to $T_s = 5 \times 10^{-3}$s. As a result, the time length after channel sensing in a slot is $T_e = 95 \times 10^{-3}$s. The mini-slot length is set to $\tau = 2 \times 10^{-3}$s and the access probability is set to $p_a = 0.3$. In addition, the step size of the learning algorithm is set to $b = 0.15$.

The simulation results mainly include the following two parts. In the first part, we illustrate the convergence of the algorithm. In the second part, we evaluate the throughput and fairness performance of the algorithm. Specifically, we compare the achievable system throughput and fairness of the following three channel selection schemes: (i) the proposed learning algorithm, (ii) the exhaustive search and (iii) the random selection approach. In the exhaustive search, it is assumed that there is an omnipotent controller which knows all the system parameters including the channel availability statistics, θ_m, and the number of secondary users, N. Based on this assumption, the exhaustive search is implemented by
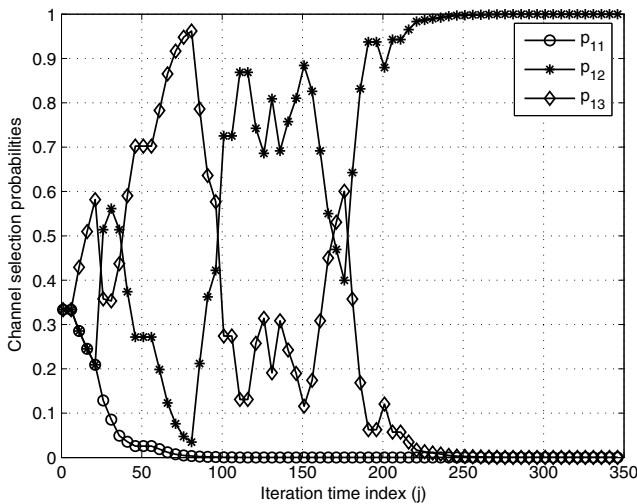
Fig. 2. Evolution of the channel selection probability of an arbitrarily secondary user ($N = 6$, $R_1 = 2$, $R_2 = 1.5$, $R_3 = 1$, $\theta_1 = 0.6$, $\theta_2 = 0.7$, $\theta_3 = 0.6$).
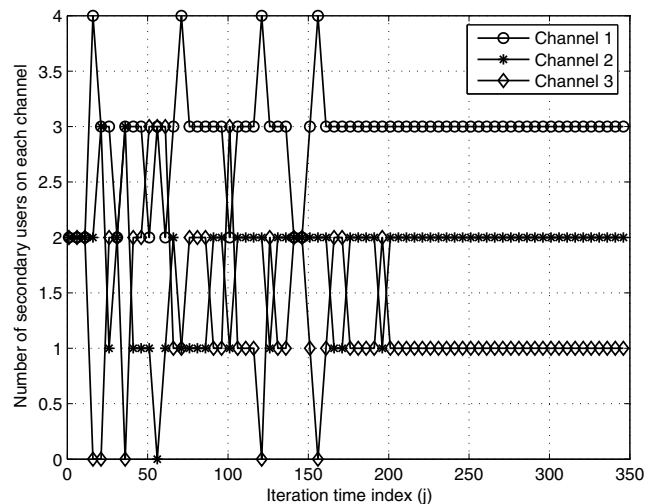

Fig. 3. Evolution of the number of secondary users selecting each channel ($N = 6$, $R_1 = 2$, $R_2 = 1.5$, $R_3 = 1$, $\theta_1 = 0.6$, $\theta_2 = 0.7$, $\theta_3 = 0.6$).
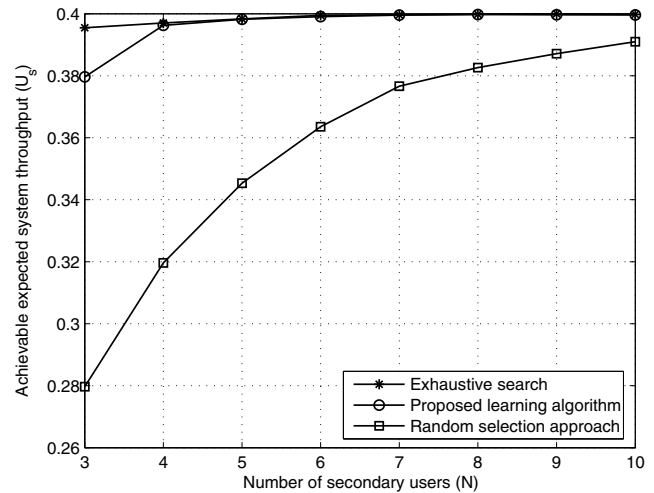

Fig. 4. Comparison of the achievable system throughput of three channel selection schemes in the homogeneous OSA system ($\theta_1 = \theta_2 = \theta_3 = 0.6$).

the controller in a centralized manner. It is clear that the achievable system throughput of the exhaustive search serves as an upper bound. On the other hand, in the random selection scheme, each secondary user selects an arbitrary licensed channel to sense with equal probability in each slot. Such a random scheme is an intuitive simple method in an OSA system where the channel availability statistics are apriori unknown and there is no information exchange.

A. Convergence of the proposed SLA based learning algorithm

First, we show the evolution of the channel selection probabilities of the SLA based channel selection algorithm. We consider an OSA system with six secondary users and three licensed channels. The channel rates and idle probabilities are set as follows: $R_1 = 2$, $R_2 = 1.5$, $R_3 = 1$, and $\theta_1 = 0.6$, $\theta_2 = 0.7$, $\theta_3 = 0.6$. Applying the genie-aided algorithm, it is known that the channel selection profile of NE solution in this case is determined by:

$$s_1 = 3, s_2 = 2, s_3 = 1. \tag{28}$$

That is, there are three secondary users selecting channel-1, two secondary users selecting channel-2 and one secondary user selecting channel-3 in the NE solution.

We plot the evolution of the channel selection probability of the proposed learning algorithm for an arbitrarily secondary user in Fig. 2. It is noted that the channel selection probability vector evolves from $\{1/3, 1/3, 1/3\}$ to $\{0, 1, 0\}$ in about 250 iterations. In other words, the secondary user finally selects channel-2 for transmission. Moreover, the evolution of the number of secondary users selecting each channel is shown in Fig. 3. It is noted that when it converges, the channel selection result is $s_1^* = 3, s_2^* = 2, s_3^* = 1$, which is exactly the NE solution of $\mathcal{G}_c$.

We also simulated more complex scenarios involving more secondary users and more licensed channels under different system parameters. The simulation results show that the

proposed SLA based algorithm always converges to a pure strategy NE point for the game.

B. Throughput performance of a homogeneous OSA system

In this subsection, we consider a homogeneous OSA system, where the licensed channels have the same transmission rates and the same availability statistics. The system parameters are set as follows: $R_1 = R_2 = R_3 = 1$, and $\theta_1 = \theta_2 = \theta_3 = 0.6$. Fig. 4 shows the comparison results of the achievable system throughput when increasing the number of secondary users. For the proposed learning algorithm and random selection approach, the simulation results are obtained by independently simulating 10^5 trials and then taking the average results. It is noted that the proposed learning algorithm significantly outperforms the random selection approach. The reasons are as follows: (i) when the proposed learning algorithm converges towards a pure strategy NE point, all the secondary users finally are spread over different channels, whereas (ii) with the random selection approach, there is a certain probability that some secondary users are crowded on
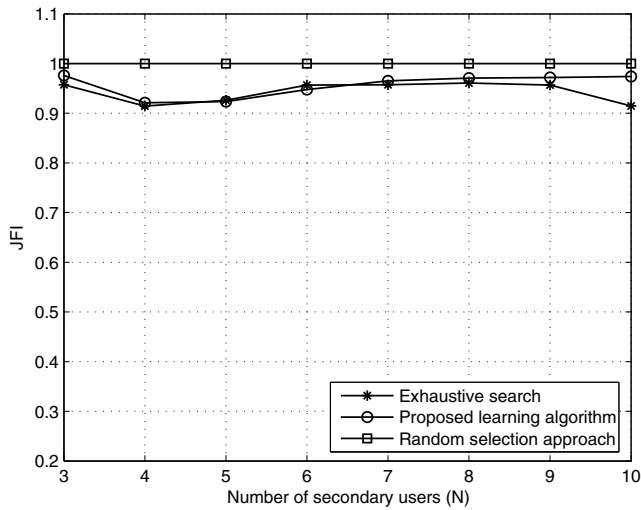
Fig. 5. Comparison results of the JFI of three channel selection schemes in the homogeneous OSA system ($\theta_1 = \theta_2 = \theta_3 = 0.6$).
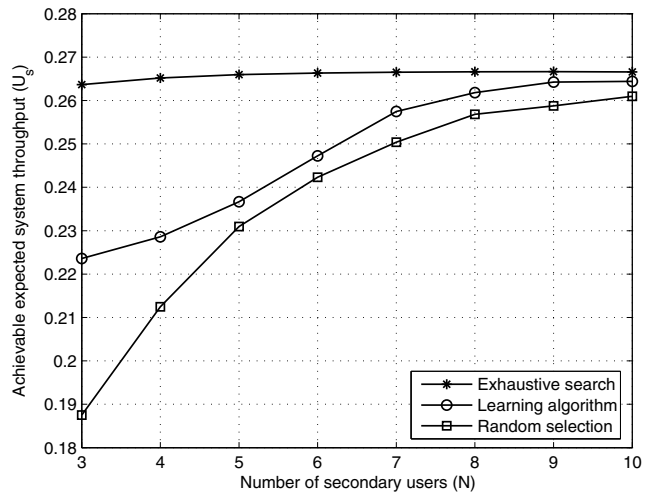

Fig. 6. Comparison of the achievable system throughput of three channel selection schemes in the heterogeneous system ($\theta_1 = 0.2, \theta_2 = 0.4, \theta_3 = 0.8$).
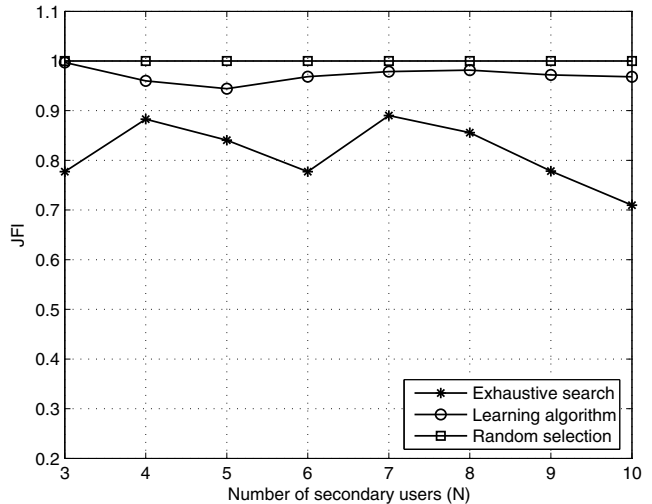

Fig. 7. Comparison results of the JFI of three channel selection schemes in the heterogeneous system ($\theta_1 = 0.2, \theta_2 = 0.4, \theta_3 = 0.8$).

a licensed channel while the other licensed channels are not selected by any secondary user. It is also noted that when the number of secondary users becomes large, e.g., $N \geq 4$, the achievable system throughput of the learning algorithm is close to that of the exhaustive search. In addition, it is noted that the throughput gap between the random selection and the exhaustive search decreases when the number of the secondary users N increases. The reason is that when N becomes sufficiently larger, the secondary users can uniformly be spread over the channels.

Moreover, the JFI of the three channel selection schemes are shown in Fig. 5, where J_e, $J_{\mathcal{G}_c}$ and J_r denote the JFI of the exhaustive search, the proposed learning algorithm and the random selection approach respectively. It is noted that the random selection approach achieves perfect fairness ($J_r \approx 1$), while the proposed learning algorithm and the exhaustive search can also achieve good fairness (both $J_{\mathcal{G}_c}$ and J_e are greater than 0.90).

From Fig. 4 and Fig. 5, it can be realized that the proposed learning algorithm can achieve the near-optimal system throughput while guaranteeing good fairness. Furthermore, the proposed learning algorithm is robust with increasing the number of secondary users. Thus, we claim that the proposed learning algorithm is efficient in the OSA system where the channel availability statistics are prior unknown and there is no information exchange among the secondary users.

C. Throughput performance of a heterogeneous OSA system

In this subsection, we consider a heterogeneous OSA system, where the variation of the licensed channels is large. The system parameters are set as follows: $R_1 = R_2 = R_3 = 1$, and $\theta_1 = 0.2, \theta_2 = 0.4, \theta_3 = 0.8$. Fig. 6 shows the comparison results of the achievable system throughput when increasing the number of secondary users. It is noted that the achievable system throughput of the proposed learning algorithm is greater than that of the random selection approach. Moreover, as the number of the secondary users N increases, the proposed learning algorithm and exhaustive search algorithms perform closely. The throughput gap between the proposed learning

algorithm and the exhaustive search is due to the nature of the heterogeneous channels. This phenomena is referred to as the *price of anarchy*, which is the inherent limitation of the NE solution, and has been well discussed in [31]–[33].

The JFI of the three channel selection schemes is shown in Fig. 7. It is noted in this case that, the random selection approach still achieves perfect fairness as expected. However, it is noted that the exhaustive search achieves poor fairness (J_e is about 0.8) whereas the proposed learning algorithm still achieves good fairness ($J_{\mathcal{G}_c}$ is greater than 0.95).

From Fig. 6 and Fig. 7, it can be realized that although the achievable system throughput of the proposed learning algorithm is slightly less than that of the exhaustive search when N is large, the learning algorithm can achieve good fairness. Thus, we claim that the learning algorithm is also desirable in this case.

TABLE I
IMPACT OF THE VARIATION OF CHANNEL AVAILABILITY STATISTICS ON
THE ACHIEVABLE SYSTEM THROUGHPUT.

Channel idle probability	Exhaustive	Learning	Random
$[0.4, 0.5, 0.5, 0.6]$	0.3797	0.3797	0.3272
$[0.25, 0.35, 0.65, 0.75]$	0.3800	0.3600	0.3248
$[0.2, 0.3, 0.6, 0.9]$	0.3801	0.3413	0.3299
$[0.15, 0.25, 0.75, 0.85]$	0.3802	0.3427	0.3269

TABLE II
IMPACT OF THE VARIATION OF CHANNEL AVAILABILITY STATISTICS ON
JFI.

Channel idle probability	Exhaustive	Learning	Random
$[0.4, 0.5, 0.5, 0.6]$	0.9705	0.9532	0.9997
$[0.25, 0.35, 0.65, 0.75]$	0.9275	0.9717	0.9989
$[0.2, 0.3, 0.6, 0.9]$	0.8481	0.9777	0.9998
$[0.15, 0.25, 0.75, 0.85]$	0.8172	0.9933	0.9994

D. Impact of the variation in channel availability statistics

It can be seen from Fig. 4 to Fig. 7 that the proposed learning algorithm leads to higher system throughput in the homogeneous OSA system than that obtained in the heterogeneous system. Thus, it is seen that the achieved performance is dependent on the system parameters. In this subsection, we investigate the impact of the variation of the channel availability statistics on the performance of the proposed learning algorithm.

We consider an OSA system involving four licensed channels and seven secondary users. For the purpose of comparison, the channel transmission rates are set equally, i.e., $R_1 = R_2 = R_3 = R_4 = 1$, but the channel availability statistics vary slightly or sharply for different scenarios. Table I and Table II show the comparison results of the achieved system throughput and the JFI in different scenarios, respectively. It is noted from Table I that the smaller the variation of the licensed channels statistics, the higher system throughput can be achieved of the proposed learning algorithm. In comparison, it is noted from Table II that when the channel availability statistics vary sharply, e.g., $[\theta_1, \ldots, \theta_4] = [0.2, 0.3, 0.6, 0.9]$, $J_{\mathcal{G}_c}$ is significantly greater than J_e. Thus, the proposed algorithm can still achieve good fairness whereas the exhaustive search achieves poor fairness in this scenario. Thus, in order to achieve high system throughput, sharply heterogeneous channels should be avoided and the quasi homogeneous channels are preferred. In practice, the techniques of channel fragmentation or channel bonding can be used to achieve quasi homogeneous channels [31], [34], [35].

VII. CONCLUSION

We studied the problem of distributed channel selection in an opportunistic spectrum access (OSA) system with unknown channel availability statistics and the number of secondary users, using a game theoretic learning solution. We formulated the channel selection problem as a game with random reward

and proved that it is an exact potential game. Firstly, we proposed a genie-aided algorithm to investigate the properties of the considered game in terms the throughput performance and the achieved fairness. Due to the lack of information and the dynamics of the environment, Nash equilibrium (NE) points of the game can not be achieved by existing methods. We then proposed a stochastic learning automata (SLA) based algorithm, with which the secondary users learn from their individual action-reward histories and adjust their behaviors towards a NE point. The proposed learning algorithm neither needs information exchange, nor requires prior information about the channel availability statistics and the number of the secondary users. Moreover, it just utilizes the individual history information of each secondary user. Thus, it can be applied to other problems in OSA systems, especially the one with unknown and dynamic environment.

However, it is also seen that we have assumed that all the users have the same access probabilities (p_a) and they remain fixed across all channels, all users, and all slots. In future work, we will investigate the issue of optimizing the access probabilities for different users, different channels and different slots, to further improve the system performance. Furthermore, it is noted that the exact potential game formulation in this work essentially relies on the fact that the users have symmetrical rates in accessing a channel. We will also consider the scenario with asymmetrical channel rates in the near future.

APPENDIX A
PROOF OF THEOREM 2

After the kth iteration, $1 \leq k < N - 1$, let us denote the set of the secondary users which have been scheduled on channel m as $c_m(k)$, i.e., $c_m(k) = \{n \in \mathcal{N}_2 : a_n = m\}$. Now, suppose that an arbitrary secondary user n is selected from $\mathcal{N}_1$ and is scheduled to transmit on channel m_0, i.e., $a_n = m_0$. As a result, we have $s_{m_0}(k + 1) = s_{m_0}(k) + 1$ and $s_m(k+1) = s_m(k), \forall m \neq m_0$. According to the channel selection principle specified by (16), it is clear that the utility function of any secondary user $i \in c_{m_0}(k + 1)$ is given by:

$$u_i = \frac{\theta_{m_0} f(s_{m_0}(k+1)) R_{m_0}}{s_{m_0}(k+1)} \geq \frac{\theta_m f(s_m(k+1)+1) R_m}{s_m(k+1)+1},$$
(A.29)

where $1 \leq m \leq M$ and $m \neq m_0$. It can be seen from (A.29) that every secondary user i, $\forall i \in c_{m_0}(k + 1)$, is not willing to deviate unilaterally.

Now, consider the utility function of an arbitrary secondary user j, $j \in c_m(k+1), \forall m \neq m_0$. Suppose that $s_m(k+1), m \neq m_0$ remains unchanged for k_0 iterations, $1 \leq k_0 \leq k$, i.e., $s_m(k + 1) = s_m(k + 1 - k_0)$. Thus, the utility function of each secondary user j, $j \in c_m(k+1), \forall m \neq m_0$, is given by:

$$u_j = \frac{\theta_m f(s_m(k+1)) R_m}{s_m(k+1)} = \frac{\theta_m f(s_m(k+1-k_0)) R_m}{s_m(k+1-k_0)}$$
$$\geq \frac{\theta_{\hat{m}} f(s_{\hat{m}}(k+1-k_0)+1) R_{\hat{m}}}{s_{\hat{m}}(k+1-k_0)+1} \geq \frac{\theta_{\hat{m}} f(s_{\hat{m}}(k+1)+1) R_{\hat{m}}}{s_{\hat{m}}(k+1)+1}, \forall \hat{m} \neq m,$$
(A.30)

where $\frac{\theta_m f(s_m(k+1-k_0))R_m}{s_m(k+1-k_0)} \geq \frac{\theta_{\hat{m}} f(s_{\hat{m}}(k+1-k_0)+1)R_{\hat{m}}}{s_{\hat{m}}(k+1-k_0)+1}$ is due to the result of the $(k-k_0)$th channel selection and $\frac{\theta_{\hat{m}} f(s_{\hat{m}}(k+1-k_0)+1)R_{\hat{m}}}{s_{\hat{m}}(k+1-k_0)+1} \geq \frac{\theta_{\hat{m}} f(s_{\hat{m}}(k+1)+1)R_{\hat{m}}}{s_{\hat{m}}(k+1)+1}$ is caused by the following fact: (1) According to Remark 1, it is known that $f(s_m)/s_m$ is a decreasing function of s_m, and (2) the inequality $s_{\hat{m}}(k+1) \geq s_{\hat{m}}(k+1-k_0)$ always holds for $\forall \hat{m} \neq m$. It can be seen from (A.30) that every secondary user $j \in c_m(k+1), \forall m \neq m_0$ is not willing to deviate unilaterally.

In conclusion, it is seen from (A.29) and (A.30) that the arbitrarily selected secondary user n in the $(k+1)$th iteration can not improve its utility function by deviating unilaterally and so do all the previously selected secondary users. Then, by the definition of NE, it is clear that the channel selection profile after the $(k+1)$th iteration is a pure strategy NE point. Inductively, after the Nth iteration, i.e., when all the secondary users are finally selected, the final channel selection scheme $\{s_1(N), \ldots, s_M(N)\}$ is a pure strategy NE point of $\mathcal{G}_c$.

APPENDIX B
PROOF OF THEOREM 4

Denote $a^* = (a_1^*, \ldots, a_N^*)$ as a pure strategy NE of $\mathcal{G}_c$, and u_n^* as the corresponding utility function of each player n. For any arbitrarily user n, we have:

$$u_n^* \geq \left(\frac{\theta_{a_k^*} f(s_{a_k^*})R_{a_k^*}}{s_{a_k^*}} \right) \left(\frac{f(s_{a_k^*}+1)}{f(s_{a_k^*})} \right) \left(\frac{s_{a_k^*}}{s_{a_k^*}+1} \right) \quad \forall k \neq n, \tag{B.31}$$

which can straightforwardly be obtained from the definition of NE specified by (11).

Since the channel contention overhead is assumed to be negligible, following Remark 1, we have $f(s_{a_k^*}+1) = f(s_{a_k^*}) = 1$. Thus, the following can be obtained:

$$u_n^* \geq u_k^* \left(\frac{s_{a_k^*}}{s_{a_k^*}+1} \right) \quad \forall n \in \mathcal{N}, k \neq n, \tag{B.32}$$

where we use the fact that $u_k^* = \frac{\theta_{a_k^*} f(s_{a_k^*})R_{a_k^*}}{s_{a_k^*}}$.

Moreover, since $s_{a_n^*} \geq 1, \forall n \in \mathcal{N}$, (B.32) can be re-written as follows:

$$u_n^* \geq \frac{1}{2} u_k^* \quad \forall n \in \mathcal{N}, k \neq n. \tag{B.33}$$

Now, let us define $u_{min}^* = \min\{u_1^*, \ldots, u_N^*\}$, $u_{max}^* = \max\{u_1^*, \ldots, u_N^*\}$ and $\rho = \frac{u_{max}^*}{u_{min}^*}$. The following inequality can be immediately obtained from (B.33):

$$1 \leq \rho \leq 2. \tag{B.34}$$

Then, according to [28], the JFI of $\mathcal{G}_c$ is bounded by:

$$J_{\mathcal{G}_c} \geq \frac{4\rho}{(\rho+1)^2}. \tag{B.35}$$

Furthermore, it is easily known that $\frac{4\rho}{(\rho+1)^2}$ is a decreasing function of ρ in the interval $[1, \infty)$. Therefore, the following lower bound of $J_{\mathcal{G}_c}$ can be obtained from (B.34) and (B.35):

$$J_{\mathcal{G}_c} \geq \frac{8}{9} \tag{B.36}$$

Thus, Theorem 4 follows.

APPENDIX C
PROOF OF THEOREM 5

We can re-write the ODE specified by (23) as follows:

$$\frac{dp_{nm}}{dt} = q F_{nm}(\mathbf{P}), \forall n \in \mathcal{N}, 1 \leq m \leq M. \tag{C.37}$$

Applying (24), (C.37) can be further re-written as:

$$\begin{aligned} \frac{dp_{nm}}{dt} &= \frac{q}{R_{max}} \Bigg(p_{nm}(1-p_{nm})\mathbf{E}[r_n|(m, \mathbf{P}_{-n})] \\ &\quad + \sum_{k=1, k\neq m}^{M} p_{nk}(-p_{nm})\mathbf{E}[r_n|(k, \mathbf{P}_{-n})] \Bigg) \\ &= \frac{qp_{nm}}{R_{max}} \sum_{k=1}^{M} p_{nk}[h_{nm}(\mathbf{P}) - h_{nk}(\mathbf{P})]. \end{aligned}$$
$$\tag{C.38}$$

It is known that the variation of $H(\mathbf{P})$ is given by:

$$\frac{\partial H(\mathbf{P})}{\partial p_{nm}} = H(m, \mathbf{P}_{-n}), \tag{C.39}$$

where we use the fact that $H(\mathbf{P}) = \sum_{m=1}^{M} p_{nm} H(m, \mathbf{P}_{-n})$.

Applying (25), (C.38) and (C.39), the derivation of $H(\mathbf{P})$ is given by:

$$\begin{aligned} \frac{dH(\mathbf{P})}{dt} &= \sum_{n,m} \frac{\partial H(\mathbf{P})}{\partial p_{nm}} \frac{dp_{nm}}{dt} \\ &= \sum_{n,m} H(m, \mathbf{P}_{-n}) \frac{qp_{nm}}{R_{max}} \sum_k p_{nk}[h_{nm}(\mathbf{P}) - h_{nk}(\mathbf{P})] \\ &= \frac{q}{R_{max}} \sum_{n,m,k} H(m, \mathbf{P}_{-n}) p_{nm} p_{nk}[h_{nm}(\mathbf{P}) - h_{nk}(\mathbf{P})] \\ &= \frac{q}{2R_{max}} \sum_{n,m,k} Y(n, m, k, \mathbf{P}) \\ &= \frac{qc}{R_{max}} \sum_{n,m,k>m} p_{nm} p_{nk}[h_{nm}(\mathbf{P}) - h_{nk}(\mathbf{P})]^2 \geq 0, \end{aligned}$$
$$\tag{C.40}$$

where $Y = p_{nm} p_{nk}[H(m, \mathbf{P}_{-n}) - H(k, \mathbf{P}_{-n})][h_{nm}(\mathbf{P}) - h_{nk}(\mathbf{P})]$. From (C.40), (C.38) and (C.37), the following is known:

$$\begin{aligned} \frac{dH(\mathbf{P})}{dt} = 0 &\Rightarrow p_{nm} p_{nk}[h_{nm}(\mathbf{P}) - h_{nk}(\mathbf{P})] = 0 \,\forall n, m, k \\ &\Rightarrow F_{nm}(\mathbf{P}) = 0 \,\, \forall n, m \\ &\Rightarrow \mathbf{P} \text{ is the stationary point of (23)} \end{aligned}$$
$$\tag{C.41}$$

In other words, the sequence $\{\mathbf{P}(j)\}$ converges to a stationary point of the ODE of (23). Thus, according to Proposition 2, Theorem 5 follows.

REFERENCES

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, 2005.

[3] C. T. Chou, N. S. Shankar, H. Kim, *et al.*, "What and how much to gain by spectrum agility?" *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 576–588, 2007.

[4] S. Srinivasa and S. Jafar, "How much spectrum sharing is optimal in cognitive radio networks?" *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 4010–4018, 2008.

[5] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Mobile Networks & Applications*, vol. 11, no. 6, pp. 779–797, 2006.

[6] M. Felegyhazi, M. Cagalj, and J. P. Hubaux, "Efficient MAC in cognitive radio systems: a game-theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1984–1995, 2009.

[7] M. Maskery, V. Krishnamurthy, and Q. Zhao, "Decentralized dynamic spectrum access for cognitive radios: cooperative design of a non-cooperative game," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 459–469, 2009.

[8] P. Sastry, V. Phansalkar, and M. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Trans. Syst., Man, Cybern. B*, vol. 24, no. 5, pp. 769–777, 1994.

[9] Q. Zhao, L. Tong, A. Swami, *et al.*, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 589–600, 2007.

[10] S. Ahmad, M. Liu, T. Javidi, *et al.*, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4040–4050, 2009.

[11] Y. Chen, Q. Zhao, and A. Swami, "Distributed spectrum sensing and access in cognitive radio networks with energy constraint," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 783–797, 2009.

[12] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.

[13] K. Liu and Q. Zhao, "Decentralized multi-armed bandit with multiple distributed players," in *Proc. 2010 Information Theory and Applications Workshop*, pp. 1–10.

[14] Y. Gai, B. Krishnamachari, and R. Jain, "Learning multiuser channel allocations in cognitive radio networks: a combinatorial multi-armed bandit formulation," in *Proc. IEEE DySPAN 2010*, pp. 1–9.

[15] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, 2010.

[16] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: learning under competition," in *Proc. 2010 IEEE INFOCOM*, pp. 1–9.

[17] H. Li, "Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: a two by two case," in *Proc. 2009 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1893–1898.

[18] Y. Xu, Q. Wu, and J. Wang, "Game theoretic channel selection for opportunistic spectrum access with unknown prior information," in *Proc. IEEE ICC 2011*, pp. 1–5.

[19] T. Joshi, D. Ahuja, D. Singh, *et al.*, "SARA: stochastic automata rate adaptation for IEEE 802.11 networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 19, no. 11, pp. 1579–1590, 2008.

[20] Y. Xing and R. Chandramouli, "Stochastic learning solution for distributed discrete power control game in wireless data networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 932–944, 2008.

[21] W. Zhong, Y. Xu, M. Tao, *et al.*, "Game theoretic multimode precoding strategy selection for MIMO multiple access channels," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 563–566, 2010.

[22] Y. Xu, J. Wang, Q. Wu, *et al.*, "Opportunistic spectrum access in cognitive radio networks: global optimization using local interaction games," *IEEE J. Sel. Topics Signal Process.* doi: 10.1109/JSTSP.2011.2176916.

[23] IEEE 802.16e-2005 and IEEE Std 802.16-2004/Cor1-2005, http://www.ieee802.org/16/.

[24] T. Cui, L. Chen, and S. H. Low, "A game-theoretic framework for medium access control," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1116–1127, 2008.

[25] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 205–222, 2009.

[26] B. Vcking and R. Aachen, "Congestion games: optimization in competition," in *Proc. 2006 Algorithms and Complexity in Durham Workshop*, pp. 9–20.

[27] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[28] R. Jain, D. Chiu, and W. Haws, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," *Technical Report*, 1984.

[29] Y. Xu, Q. Wu, J. Wang, *et al.*, "Social welfare maximization for SRSNs using bio-inspired community cooperation mechanism," *Chinese Sci. Bull.*, vol. 57, no. 1, pp. 125–131, 2012.

[30] J. Marden, G. Arslan, and J. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Trans. Automatic Control*, vol. 54, no. 2, pp. 208–220, 2009.

[31] L. Law, J. Huang, M. Liu, and S. R. Li, "Price of anarchy for cognitive MAC games," in *Proc. IEEE GLOBECOM 2009*, pp. 1–6.

[32] P. Chaporkar, A. Proutiere, and B. Radunoviac, "Rate adaptation games in wireless LANs: Nash equilibrium and price of anarchy," in *Proc. IEEE INFOCOM 2010*, pp. 1–9.

[33] U. Ayesta, O. Brun, and B. J. Prabhu, "Price of anarchy in non-cooperative load balancing," in *Proc. IEEE INFOCOM 2010*, pp. 1–5.

[34] L. Jiao, V. Pla, and F. Li, "Analysis on channel bonding/aggregation for multi-channel cognitive radio networks," in *Proc. 2010 European Wireless Conference*, pp. 468–474.

[35] Y. Lu, H. He, J. Wang, and S. Li, "Opportunistic spectrum access with channel bonding," in *Proc. 2009 International Conference on Communications and Networking in China*, pp. 1–5, 2009.

Yuhua Xu received his B.S. degree in communications engineering, M.S. degree in communications and information system from Institute of Communications Engineering, Nanjing, China, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree in communications and information system in Institute of Communications Engineering, PLA University of Science and Technology. His research interests focus on opportunistic spectrum access, learning theory, game theory, and optimization techniques.

Jinlong Wang received his B.S. degree in wireless communications, M.S. degree and Ph.D. degree in communications and electronic system from Institute of Communications Engineering, Nanjing, China, in 1983, 1986 and 1992, respectively. He is currently a professor at the PLA University of Science and Technology, China. He is also the co-chairman of IEEE Nanjing Section. He has published widely in the areas of signal processing for communications, information theory, and wireless networks. His current research interests include wireless communication, cognitive radio, soft-defined radio and ultra-wide bandwidth (UWB) systems.

Qihui Wu received his B.S. degree in communications engineering, M.S. degree and Ph.D. degree in communications and information system from Institute of Communications Engineering, Nanjing, China, in 1994, 1997 and 2000, respectively. He is currently a professor at the PLA University of Science and Technology, China. His current research interests are algorithms and optimization for cognitive wireless networks, soft-defined radio and wireless communication systems.

Alagan Anpalagan received the B.A.Sc. (H), M.A.Sc. and Ph.D. degrees in Electrical Engineering from the University of Toronto, Canada. He joined the ELCE Department of at Ryerson University in 2001 and was promoted to Full Professor in 2010. He served the department as Graduate Program Director (2004-09) and the Interim Electrical Engineering Program Director (2009-10). During his sabbatical (2010-11), he was a Visiting Professor at Asian Institute of Technology and Visiting Researcher at Kyoto University. Dr. Anpalagan's industrial experience includes working at Bell Mobility on 1xRTT system deployment studies (2001), at Nortel Networks on SECORE R&D projects (1997) and at IBM Canada as IIP Intern (1994).

Dr. Anpalagan directs a research group working on radio resource management (RRM) and radio access & networking (RAN) areas within the WINCORE Lab that mainly focuses on cross layer design, analysis and optimization of wireless systems. His current research interests include cognitive radio RRM, wireless cross layer design and optimization, collaborative communication, green communications technologies and QoE-aware femtocells.

Dr. Anpalagan serves as Associate Editor for the IEEE COMMUNICA-TIONS LETTERS (2010-12) and *Springer Wireless Personal Communications* (2009-12), and past Editor for *EURASIP Journal of Wireless Communications and Networking* (2004-2009). He also served as EURASIP Guest Editor for two special issues in RRM in 3G+ Systems (2006) and Fairness in RRM for Wireless Networks (2008). Dr. Anpalagan served as TPC Co-Chair of: IEEE PIMRC'11 Track on Cognitive Radio and Spectrum Management, IEEE IWCMC'11 Workshop on Cooperative and Cognitive Networks, IEEE CCECE'04/'08 and WirelessCom'05 Symposium on RRM. Dr. Anpalagan served as IEEE Toronto Section Chair (2006-07), ComSoc Toronto Chapter Chair (2004-05), Chair of IEEE Canada Professional Activities Committee (2009-11). He is the recipient of the Dean's Teaching Award (2011), Faculty Scholastic, Research and Creativity Award (2010), Faculty Service Award (2010) at Ryerson University. Dr. Anpalagan also completed a course on Project Management for Scientist and Engineers at the University of Oxford CPD Center. He is a registered Professional Engineer in the province of Ontario, Canada.

Yu-Dong Yao (S88-M88-SM94) has been with Stevens Institute of Technology, Hoboken, New Jersey, since 2000 and is currently a professor and department director of electrical and computer engineering. He is also a director of Stevens' Wireless Information Systems Engineering Laboratory (WISELAB). Previously, from 1989 and 1990, he was at Carleton University, Ottawa, Canada, as a Research Associate working on mobile radio communications. From 1990 to 1994, he was with Spar Aerospace Ltd., Montreal, Canada, where he was involved in research on satellite communications. From 1994 to 2000, he was with Qualcomm Inc., San Diego, CA, where he participated in research and development in wireless code-division multiple-access (CDMA) systems.

He holds one Chinese patent and twelve U.S. patents. His research interests include wireless communications and networks, spread spectrum and CDMA, antenna arrays and beamforming, cognitive and software defined radio (CSDR), and digital signal processing for wireless systems. Dr. Yao was an Associate Editor of IEEE COMMUNICATIONS LETTERS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He received the B.Eng. and M.Eng. degrees from Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Southeast University, Nanjing, China, in 1988, all in electrical engineering.